

# Metastable distributions of Markov chains with rare transitions

M. Freidlin\*, L. Koralov†

## Abstract

In this paper we consider Markov chains with transition rates that depend on a small parameter  $\varepsilon$ . Under a mild assumption on the asymptotics of these transition rates, we describe the behavior of the chain at various  $\varepsilon$ -dependent time scales, i.e., we calculate the metastable distributions.

2010 Mathematics Subject Classification Numbers: 37A25, 37A30, 60J27, 60J28, 60F10.

Keywords: Markov Chains, Metastable Distributions, Large Deviations.

## 1 Introduction.

Consider a family  $X_t^\varepsilon$  of Markov chains on a state space  $S = \{1, \dots, N\}$ , where  $\varepsilon$  is a small parameter. The time may be continuous or discrete – we start by considering the case when  $t \in \mathbb{R}^+$ . The case of discrete time is similar and is briefly discussed in Section 6. Let  $q_{ij}(\varepsilon)$ ,  $i, j \in S$ ,  $i \neq j$ , be the transition rates, i.e.,

$$P(X_{t+\Delta}^\varepsilon = j | X_t^\varepsilon = i) = q_{ij}(\varepsilon)(\Delta + o(\Delta)) \quad \text{as } \Delta \downarrow 0, \quad i \neq j. \quad (1)$$

We will be interested in the behavior of  $X_t^\varepsilon$  as  $\varepsilon \downarrow 0$  and, simultaneously,  $t = t(\varepsilon) \rightarrow \infty$ . The results on the asymptotic behavior of  $X_t^\varepsilon$  can be viewed as a refinement of the ergodic theorem for Markov chains (which concerns the asymptotics with respect to the time variable only) and are obviously closely related to the spectral properties of the transition matrix. The double limit at hand depends on how the point  $(1/\varepsilon, t(\varepsilon))$  approaches infinity. Roughly speaking, one can divide the neighborhood of infinity into a finite number of domains such that  $X_{t(\varepsilon)}^\varepsilon$  has a limiting distribution (which depends on the initial point) when  $(1/\varepsilon, t(\varepsilon))$  approaches infinity without leaving a given domain. For different domains, these limits are different. These will be referred to as metastable distributions.

---

\*Dept of Mathematics, University of Maryland, College Park, MD 20742, mif@math.umd.edu

†Dept of Mathematics, University of Maryland, College Park, MD 20742, koralov@math.umd.edu

Families of parameter-dependent Markov chains arise in a variety of applications. In particular, this is a natural object in problems concerning random perturbations of dynamical systems ([5]). If a dynamical system has  $N$  asymptotically stable attractors, each attractor (or rather its small neighborhood) can be associated with a state of a Markov chain, while the transition times between different states are due to large deviations and are determined by the action functional of the perturbed system. In this example, these transition times are exponentially large, i.e., the transition rates are exponentially small with respect to the size of the perturbation  $\varepsilon$ .

As shown in [3], [5], the long-time behavior of the perturbed process can typically be understood using the notion of the hierarchy of cycles. The hierarchy of cycles means, roughly speaking, that for each  $0 \leq r \leq \rho$  the set of attractors is decomposed into disjoint subsets  $C_1^r, \dots, C_{n_r}^r$  (cycles of rank  $r$ ), up to the maximal rank  $\rho < N$ . The individual attractors are the cycles of rank zero, they are combined in disjoint sets - cycles of rank one, those are combined in cycles of rank two, etc., until the cycle of maximal rank  $\rho < N$  containing all the attractors. With probability close to one as  $\varepsilon \downarrow 0$ , the process goes from a neighborhood of one of the attractors in  $C_i^r$  to a small neighborhood of one of the attractors of the next cycle  $C_j^{r+1}$ , thus remaining within a cycle of rank  $r+1$ . The transition time between  $C_i^r$  and  $C_j^{r+1}$  is determined by the asymptotics, as  $\varepsilon \downarrow 0$ , of the transition rates between individual attractors that belong to the union of these cycles. The process leaves the cycle of rank  $r+1$  only after an exponentially large number of transitions between cycles of rank  $r$ .

For each  $\lambda$  (except a finite number of values) and the time scale  $t(\varepsilon) \sim \exp(\lambda/\varepsilon)$ , with probability close to one, the process can be found in a neighborhood of a particular attractor (meta-stable state), which depends on the initial state. The meta-stable state is a piece-wise constant function of the parameter  $\lambda$ .

The above description with the hierarchy of cycles and the meta-stable states is valid, however, only if the notion of the unique “next” cycle (and, consequently, unique meta-stable state) can be correctly defined, which is not the case in many interesting situations. For example, when the unperturbed dynamics has certain symmetries (or “rough symmetries” as in [4]), the notion of “next” is not defined uniquely even for individual attractors (cycles of rank zero). A similar phenomenon was observed in [6], due not to symmetries but to degeneration of the unperturbed dynamics in a part of the phase space. Other systems leading to parameter-dependent Markov chains with no unique metastable state arise in the study of various models of non-equilibrium statistical mechanics at low temperatures (see [8] and references therein). The transition rates for such chains typically decay exponentially in the parameter corresponding to inverse temperature. Various large deviation results for Markov chains with exponentially small transition rates were obtained in [11] [9], [10].

Yet another example is provided by dynamical systems with heteroclinic networks. Namely, assume that there are finitely many stationary points with heteroclinic connections that together form a connected set. Assume that the entire network serves as an attractor for the dynamical system. The flow lines (or their sufficiently small neighborhoods) connecting pairs of stationary points can be associated with the states of a Markov

chain. After a random perturbation of size  $\varepsilon$ , the transition times between the neighborhoods of such flow lines scale as powers of  $\varepsilon$  (up to a logarithmic factor). So the notions of the hierarchy of cycles and the meta-stable states could apply (at times that scale as powers of  $\varepsilon$ , rather than exponentially). However, the notion of “next” state may again be not defined uniquely, since the exit from a neighborhood of a heteroclinic connection (say, between stationary points  $A$  and  $B$ ) can happen along either of the heteroclinic connections leading out of  $B$ . A detailed study of motion along heteroclinic networks is the subject of [1]. It should be stressed that the dynamics in this example, as well as in the case of asymptotically stable attractors discussed above, is only approximately described by a Markov chain of the type considered in the current paper. A reduction of the true dynamics to a finite-state Markov chain requires non-trivial analysis.

In the current paper, we introduce the notion of the hierarchy of Markov chains in a general setting. We show that it should replace the notion of the hierarchy of cycles. The meta-stable states are replaced by meta-stable distributions. We do not require the transition rates between different states to scale exponentially (or have any specific asymptotic behavior), but only assume that there is a certain asymptotic relation between the ratios of transition rates.

More precisely, we will say that the family  $X_t^\varepsilon$  is *asymptotically regular* as  $\varepsilon \downarrow 0$  if the following two conditions hold.

- (a) The transition rates  $q_{ij}(\varepsilon)$  are positive<sup>1</sup> for  $\varepsilon > 0$  and all  $i \neq j$ .
- (b) For each  $i, j, k, l \in S$  satisfying  $i \neq j$ ,  $k \neq l$ , the following finite or infinite limit exists

$$\lim_{\varepsilon \downarrow 0} (q_{ij}(\varepsilon)/q_{kl}(\varepsilon)) \in [0, \infty].$$

In Section 2, we will introduce the hierarchy of Markov chains. The hierarchy will be defined inductively by successively reducing the state space, i.e., combining the elements of the state space into subsets that will serve as states for the chain of higher rank. In order to perform an inductive step, we will require each chain appearing in the construction to be asymptotically regular. While this condition may seem not quite explicit, we will show that it holds if the transition rates of the original family satisfy a relatively simple condition.

In Section 3, we discuss some simple properties of asymptotically regular Markov chains. In Section 4, we formulate and prove the main theorem on the meta-stable behavior of the original process. In Section 5, we prove that a condition on the transition rates of the family of Markov chains guarantees that all the chains in the hierarchy are asymptotically regular. We briefly discuss a couple of generalizations in Section 6.

---

<sup>1</sup>The main result can be obtained even if the positivity assumption is relaxed, as mentioned in Section 6.

## 2 Hierarchy of Markov chains.

### 2.1 Reduced Markov chain.

Given an asymptotically regular family of Markov chains  $X_t^\varepsilon$  with  $N \geq 2$ , we will construct a reduced Markov chain (later also referred to as the reduced Markov chain of rank one). First, we define a discrete-time Markov chain on  $S$ , which will be referred to as the *skeleton Markov chain* and denoted by  $Z_n$ . Its transition probabilities are defined by

$$P_{ij} = \lim_{\varepsilon \downarrow 0} (q_{ij}(\varepsilon) / \sum_{j' \neq i} q_{ij'}(\varepsilon)), \quad j \neq i; \quad P_{ii} = 0.$$

Observe that the above limit exists since the family of chains  $X_t^\varepsilon$  is asymptotically regular. Recall (see [2]) that the (finite) state space of a Markov chain can be uniquely decomposed into a disjoint union of ergodic classes and the sets consisting of individual transient states,

$$S = S_1 \bigcup \dots \bigcup S_n. \quad (2)$$

Note that  $n < N$  since each ergodic class of  $Z_n$  contains at least two states, which follows from the fact that for each  $i$  there is  $j \neq i$  such that  $P_{ij} > 0$ .

Next, for each  $1 \leq k \leq n$ , we define the Markov chains  $Y_t^{k,\varepsilon}$  by narrowing the state space  $S$  to  $S_k$  and retaining the same transition rates for  $i, j \in S_k$  as in the original Markov chain  $X_t^\varepsilon$ . Let  $\mu^k(i, \varepsilon)$  be the invariant measure of the state  $i \in S_k$  for the chain  $Y_t^{k,\varepsilon}$ .

Finally, we define the reduced Markov chain. Its state space is the set  $\{S_1, \dots, S_n\}$ . The transition rate between  $S_k$  and  $S_l$ ,  $k \neq l$ , denoted by  $Q_{kl}(\varepsilon)$ , is defined by

$$Q_{kl}(\varepsilon) = \sum_{i \in S_k} \sum_{j \in S_l} \mu^k(i, \varepsilon) q_{ij}(\varepsilon). \quad (3)$$

### 2.2 Definition of the hierarchy.

We will use induction to define the *reduced Markov chains*  $X_t^{r,\varepsilon}$ ,  $0 \leq r \leq \rho$ , with some  $0 \leq \rho < N$ . The reduced Markov chain of rank zero (i.e., corresponding to  $r = 0$ ) will coincide with  $X_t^\varepsilon$ , while the reduced Markov chain of rank  $\rho$  will be trivial (i.e., its state space will contain one element). The entire collection of reduced Markov chains will be referred to as the hierarchy.

For each  $0 \leq r \leq \rho - 1$ , by partitioning the state space of  $X_t^{r,\varepsilon}$  into  $n_{r+1}$  subsets (referred to as clusters), we will define Markov chains  $Y_t^{r,k,\varepsilon}$ ,  $1 \leq k \leq n_{r+1}$ . The chain  $Y_t^{r,k,\varepsilon}$  will be referred to as the  $k$ -th chain of rank  $r$ .

We set  $n_0 = N$  and  $S_1^0 = 1, \dots, S_{n_0}^0 = n_0$ . These are clusters of rank zero. The reduced Markov chain of rank zero, denoted by  $X_t^{0,\varepsilon}$ , is defined on the state space  $S^0 = \{S_1^0, \dots, S_{n_0}^0\}$  and has transition rates  $Q_{ij}^0(\varepsilon) = q_{ij}(\varepsilon)$ . Thus it coincides with the original Markov chain.

If  $N = 1$ , this results in the trivial hierarchy. If  $N \geq 2$ , we can apply the above construction of the reduced Markov chain. We set  $n_1 = n$  and use the following notation

$$S = S_1^1 \bigcup \dots \bigcup S_{n_1}^1.$$

for the decomposition of the original state space into ergodic classes and sets containing individual transient states for the skeleton chain. The sets  $S_1^1, \dots, S_{n_1}^1$  will be referred to as clusters of rank one. For each  $1 \leq k \leq n_1$ , we set  $Y_t^{0,k,\varepsilon} = Y_t^{k,\varepsilon}$ , which will be referred to as the  $k$ -th Markov chain of rank zero. Let  $\mu^{0,k}(i, \varepsilon)$  be the invariant measure of the state  $S_i^0 \in S_k^1$  for the chain  $Y_t^{0,k,\varepsilon}$  defined above.

The reduced Markov chain introduced above will also be referred to as the reduced Markov chain or rank one and denoted by  $X_t^{1,\varepsilon}$ . Its state space is the set  $\{S_1^1, \dots, S_{n_1}^1\}$ . The transition rate between  $S_k^1$  and  $S_l^1$ ,  $k \neq l$ , denoted by  $Q_{kl}^1(\varepsilon)$ , is defined, in conformance with (3), by

$$Q_{kl}^1(\varepsilon) = \sum_{i \in S_k^1} \sum_{j \in S_l^1} \mu^{0,k}(i, \varepsilon) Q_{ij}^0(\varepsilon). \quad (4)$$

If  $X_t^{1,\varepsilon}$  is asymptotically regular, we can replicate the above construction, i.e., define the skeleton chain corresponding to  $X_t^{1,\varepsilon}$  and partition  $\{S_1^1, \dots, S_{n_1}^1\}$  into clusters  $S_1^2, \dots, S_{n_2}^2$ . For each  $1 \leq k \leq n_2$ , we can define the  $k$ -th Markov chain of rank one, denoted by  $Y_t^{1,k,\varepsilon}$ . The reduced Markov chain  $X_t^{2,\varepsilon}$  of rank two is defined on the state space  $\{S_1^2, \dots, S_{n_2}^2\}$ . The construction then continues inductively, assuming that all the reduced chains are asymptotically regular.

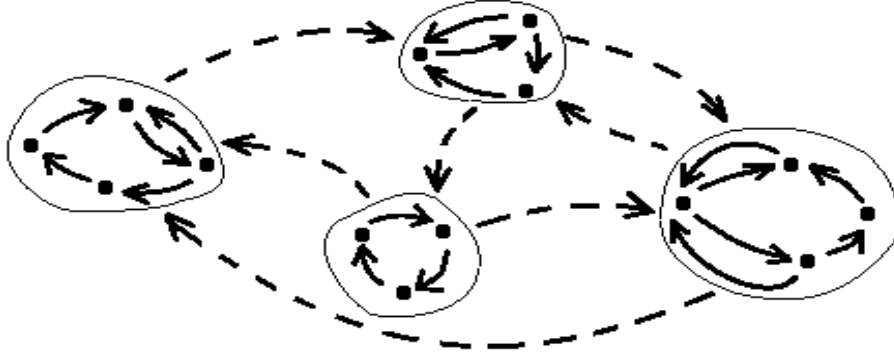


Figure 1: In this example,  $n_0 = N = 14$ ,  $n_1 = 4$ ,  $n_2 = 1$ , and  $\rho = 2$ . The solid arrows denote non-zero transitions for the skeleton Markov chain corresponding to  $X_t^{0,\varepsilon}$ . The dashed arrows denote non-zero transitions for the skeleton Markov chain corresponding to  $X_t^{1,\varepsilon}$ .

Let  $\rho$  be the first index such that  $n_\rho = 1$ . Observe that  $\rho < N$  since  $n_0 = N$  and  $n_{r+1} < n_r$  for each  $r$ . Since  $n_\rho = 1$ , there is only one Markov chain of rank  $\rho - 1$ , and the reduced Markov chain of rank  $\rho$  is trivial – its state space consists of one element  $S_1^\rho$ . This completes the construction of the hierarchy.

The reduced Markov chain of rank  $r$ ,  $0 \leq r \leq \rho$ , will be denoted by  $X_t^{r,\varepsilon}$ , its transition rates will be denoted by  $Q_{kl}^r(\varepsilon)$ , the  $k$ -th Markov chain of rank  $r$ ,  $0 \leq r < \rho$ ,  $1 \leq k \leq n_{r+1}$ , will be denoted by  $Y_t^{r,k,\varepsilon}$ , and the invariant measure of a state  $S_i^r \in S_k^{r+1}$  for this Markov chain will be denoted by  $\mu^{r,k}(i, \varepsilon)$ .

For a state  $j \in S$ , we'll write that  $j \prec S_k^r$  if there is a sequence  $j = j_0, j_1, \dots, j_r = k$  such that

$$j = S_{j_0}^0 \in S_{j_1}^1 \dots \in S_{j_r}^r. \quad (5)$$

For  $1 \leq k \leq n_r$  and  $j$  such that  $j \prec S_k^r$  does not hold, we will need the functions  $\tilde{Q}_{kj}^r(\varepsilon)$ . These are defined inductively, namely,  $\tilde{Q}_{kj}^0(\varepsilon) = Q_{ij}^0(\varepsilon)$  if  $S_k^0 = i$  and

$$\tilde{Q}_{kj}^r(\varepsilon) = \sum_{i: S_i^{r-1} \in S_k^r} \mu^{r-1,k}(i, \varepsilon) \tilde{Q}_{ij}^{r-1}(\varepsilon), \quad 1 \leq r \leq \rho - 1.$$

Intuitively, these serve as transition rates from a cluster of rank  $r$  to an individual state, although, unlike  $Q_{kl}^r(\varepsilon)$ , they don't correspond to transition rates of any of the Markov chains introduced above.

Let us stress that the inductive construction of the hierarchy is possible under the condition that all the reduced chains that appear at each step are asymptotically regular. We will say that  $X_t^\varepsilon$  is *completely asymptotically regular* if for each  $a > 0$  and each  $(i_1, \dots, i_a)$ ,  $(j_1, \dots, j_a)$ ,  $(k_1, \dots, k_a)$ , and  $(l_1, \dots, l_a)$  the following finite or infinite limit exists

$$\lim_{\varepsilon \downarrow 0} \left( \frac{q_{i_1 j_1}(\varepsilon)}{q_{k_1 l_1}(\varepsilon)} \times \dots \times \frac{q_{i_a j_a}(\varepsilon)}{q_{k_a l_a}(\varepsilon)} \right) \in [0, \infty], \quad (6)$$

provided that  $i_1 \neq j_1, \dots, i_a \neq j_a, k_1 \neq l_1, \dots, k_a \neq l_a$ .

**Lemma 2.1.** *Suppose that  $X_t^\varepsilon$  is completely asymptotically regular. Then the reduced chain is also completely asymptotically regular.*

This lemma will be proved in Section 5. For now, we observe that if  $X_t^\varepsilon$  is completely asymptotically regular, then, by Lemma 2.1, so are the reduced Markov chains that appear at each step of the inductive construction of the hierarchy, which implies that all of them are asymptotically regular.

### 3 Asymptotically regular families of Markov chains.

In order to prepare for the discussion of metastability, we need several lemmas on asymptotically regular families of Markov chains. Let  $\mu(i, \varepsilon)$  denote the invariant measure of the state  $i$  in an asymptotically regular Markov chain  $X_t^\varepsilon$  with  $N \geq 2$ . We'll be interested in the asymptotics of  $\mu(i, \varepsilon)$  in the case when the skeleton chain has one ergodic class and no transient states (in which case  $\rho = 1$  and no additional assumptions are required in

order to define the hierarchy). Note that this condition is satisfied for each of the Markov chains  $Y_t^{r,k,\varepsilon}$  defined above, with the exception that the number of states for such a chain may be equal to one.

Let  $\lambda$  be the invariant measure of the skeleton chain  $Z_n$ . Define

$$T(i, \varepsilon) = \left( \sum_{j \in S, j \neq i} q_{ij}(\varepsilon) \right)^{-1}, \quad \bar{T}(\varepsilon) = \sum_{i' \in S} (\lambda(i') T(i', \varepsilon)).$$

Thus  $T(i, \varepsilon)$  is the average of the exponentially distributed exit time of  $X_t^\varepsilon$  from the state  $i$ . The function  $\bar{T}(\varepsilon)$  is the average time it takes  $X_t^\varepsilon$  to make one step, where the average is calculated with respect to the invariant measure for the skeleton chain.

**Lemma 3.1.** *Let  $X_t^\varepsilon$  be asymptotically regular and  $N \geq 2$ . Suppose that the skeleton chain has one ergodic class and no transient states. Then*

$$\mu(i, \varepsilon) \sim \lambda(i) T(i, \varepsilon) / \bar{T}(\varepsilon) \quad \text{as } \varepsilon \downarrow 0.$$

*Proof.* Let  $t_1^\varepsilon, t_2^\varepsilon, \dots$  be the times of the jumps of the process  $X_t^\varepsilon$ . Let  $Z_n^\varepsilon$  be the discrete-time Markov chain obtained from  $X_t^\varepsilon$  by discretizing time, i.e.,  $Z_0^\varepsilon = X_0^\varepsilon$ ,  $Z_n^\varepsilon = X_{t_n}^\varepsilon$ ,  $n \geq 1$ . Let  $P_{ij}^\varepsilon$  be the transition probabilities for  $Z_n^\varepsilon$  and  $\lambda^\varepsilon$  be its invariant measure. Then, from the definition of the skeleton chain, it follows that

$$\lim_{\varepsilon \downarrow 0} P_{ij}^\varepsilon = P_{ij}, \quad i \neq j, \tag{7}$$

where  $P_{ij}$  are the transition probabilities for  $Z_n$ . Since  $Z_n$  has one ergodic class and no transient states, this implies that

$$\lim_{\varepsilon \downarrow 0} \lambda^\varepsilon(i) = \lambda(i) > 0, \quad i \in S.$$

By the Law of Large Numbers for Markov chains,  $\mu(i, \varepsilon)$  is equal to the asymptotic (as  $t \rightarrow \infty$ ) proportion of time that the process  $X_t^\varepsilon$  spends in the state  $i$ . Therefore,

$$\mu(i, \varepsilon) = \lambda^\varepsilon(i) T(i, \varepsilon) / \sum_{i' \in S} (\lambda^\varepsilon(i') T(i', \varepsilon)).$$

Combining the latter two equalities, we obtain the result claimed in the lemma.  $\square$

From the asymptotic regularity of  $X_t^\varepsilon$  it follows that there is a limiting probability measure

$$\mu(i) = \lim_{\varepsilon \downarrow 0} \mu(i, \varepsilon) \in [0, 1], \quad i \in S. \tag{8}$$

Given two functions  $t(\varepsilon), s(\varepsilon) : (0, \infty) \rightarrow (0, \infty)$ , we'll write  $s(\varepsilon) \ll t(\varepsilon)$  if  $s(\varepsilon) = o(t(\varepsilon))$  as  $\varepsilon \downarrow 0$ . Let  $\alpha(j, t)$  be the proportion of time, prior to  $t$ , that the process spends in  $j$ .

**Lemma 3.2.** *Let  $X_t^\varepsilon$  be asymptotically regular and  $N \geq 2$ . Suppose that the skeleton chain has one ergodic class and no transient states. Suppose that  $t(\varepsilon)$  is such that  $\bar{T}(\varepsilon) \ll t(\varepsilon)$ . Then for each  $i, j \in S$ ,*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = \mu(j). \quad (9)$$

$$E_i \alpha(j, t(\varepsilon)) \sim \mu(j, \varepsilon), \quad \text{as } \varepsilon \downarrow 0. \quad (10)$$

For each  $c > 0$ , there are  $\delta(c) > 0$  and  $\varepsilon_0 > 0$  such that

$$P_i(\alpha(j, t(\varepsilon)) < (1 - c)\mu(j, \varepsilon)) \leq e^{-\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0. \quad (11)$$

$$P_i(\alpha(j, t(\varepsilon)) > (1 + c)\mu(j, \varepsilon)) \leq e^{-\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0. \quad (12)$$

Here the subscript  $i$  stands for the initial location of the process.<sup>2</sup>

*Proof.* Let  $i^* \in S$  be such that  $\mu(i^*) > 0$ . Given  $i \in S$ , find a sequence  $i_0, i_1, \dots, i_k$  such that  $i_0 = i$ ,  $i_k = i^*$ ,  $0 \leq k < N$ , and  $P_{i_0 i_1}, \dots, P_{i_{k-1} i_k} > 0$ , where the latter are the transition probabilities for the skeleton Markov chain. Then, examining the transition rates of  $X_t^\varepsilon$ , it is easy to see that

$$P_i(Z_0^\varepsilon = i_0, \dots, Z_k^\varepsilon = i_k, \quad t_k^\varepsilon < \bar{T}(\varepsilon) < t_{k+1}^\varepsilon) \geq a$$

for some positive constant  $a$  and all sufficiently small  $\varepsilon$ . In particular,  $P_i(X_{\bar{T}(\varepsilon)}^\varepsilon = i^*) \geq a$ . Since  $i$  was arbitrary, this provides an upper bound on the speed of convergence of  $X_t^\varepsilon$  to the invariant distribution, i.e.,  $P_i(X_{t(\varepsilon)}^\varepsilon = j) - \mu(j, \varepsilon) \rightarrow 0$  as  $\varepsilon \downarrow 0$  if  $\bar{T}(\varepsilon) \ll t(\varepsilon)$ . Combined with (8), this implies (9).

Let  $n(\varepsilon) = t(\varepsilon)/\bar{T}(\varepsilon)$ . Let  $\beta(j, t)$  be the amount of time, prior to  $t$ , that the process spends in  $j$ . From the large deviation estimates for the Markov chains it easily follows that for  $c' \in (0, 1)$  there are  $\delta(c') > 0$  and  $\varepsilon_0 > 0$  such that

$$P_i(t_{[(1-c')n(\varepsilon)]}^\varepsilon > t(\varepsilon)) < e^{-\delta(c')n(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0. \quad (13)$$

$$P_i(t_{[(1+c')n(\varepsilon)]}^\varepsilon < t(\varepsilon)) < e^{-\delta(c')n(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0, \quad (14)$$

Moreover, for  $c' \in (0, c \wedge 1)$ , there are  $\delta(c, c') > 0$  and  $\varepsilon_0 > 0$  such that

$$P_i(\beta(j, t_{[(1-c')n(\varepsilon)]}^\varepsilon) < (1 - c)\mu(j, \varepsilon)t(\varepsilon)) < e^{-\delta(c, c')n(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0, \quad (15)$$

while for  $c' > 0$  satisfying  $(1 + c)(1 - c')/(1 + c') > 1$ , there are  $\delta(c, c') > 0$  and  $\varepsilon_0 > 0$  such that

$$P_i(\beta(j, t_{[(1+c')n(\varepsilon)]}^\varepsilon) > (1 + c)\mu(j, \varepsilon)t(\varepsilon)) < e^{-\delta(c, c')n(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0, \quad (16)$$

We obtain (11) by combining (13) and (15). We obtain (12) by combining (14) and (16).

---

<sup>2</sup>Formula (9) can be improved to  $P_i(X_{t(\varepsilon)}^\varepsilon = j) \sim \mu(j, \varepsilon)$  as  $\varepsilon \downarrow 0$ , but we don't need it here.



From (12) and the strong Markov property of the process, it follows that

$$P_i(\alpha(j, t(\varepsilon)) > k(1+c)\mu(j, \varepsilon)) \leq e^{-k\delta(c)t(\varepsilon)/\bar{T}(\varepsilon)}, \quad \varepsilon \leq \varepsilon_0,$$

for each  $k \in \mathbb{N}$ . Combined with (11), this immediately implies (10).  $\square$

Next, let us consider the behavior of an asymptotically regular chain that is stopped when it enters a non-empty set  $E \subseteq S$ . Let  $\sigma = \inf\{t : X_t^\varepsilon \in E\}$ ,  $\tau = \min\{n : Z_n \in E\}$ ,  $\tau' = \min\{n : Z_n^\varepsilon \in E\}$ .

**Lemma 3.3.** *Let  $X_t^\varepsilon$  be asymptotically regular and  $N \geq 2$ . Suppose that the skeleton chain has one ergodic class and no transient states. Let  $E$  be a non-empty subset of  $S$ . Then for each  $i \in S$  and  $j \in E$ ,*

$$\lim_{\varepsilon \downarrow 0} P_i(X_\sigma^\varepsilon = j) = P_i(Z_\tau = j). \quad (17)$$

*If  $t(\varepsilon)$  is such that  $t(\varepsilon) \ll T(i, \varepsilon)$  for  $i \in E$ ,  $T(i, \varepsilon) \ll t(\varepsilon)$  for  $i \notin E$ , then*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = P_i(Z_\tau = j). \quad (18)$$

*Proof.* Recall that  $Z_n^\varepsilon$  is the discrete-time Markov chain obtained from  $X_t^\varepsilon$  by discretizing time. Then

$$P_i(X_\sigma^\varepsilon = j) = P_i(Z_{\tau'}^\varepsilon = j) \rightarrow P_i(Z_\tau = j) \quad \text{as } \varepsilon \downarrow 0,$$

where the convergence follows from (7). Now let us prove (18). Given  $\delta > 0$ , find  $k$  such that  $P_i(\tau > k) < \delta$ . From the convergence of  $Z_n^\varepsilon$  to  $Z_n$ , it then follows that

$$P_i(\tau' > k) < \delta$$

for all sufficiently small  $\varepsilon$ . From the condition  $T(i, \varepsilon) \ll t(\varepsilon)$  for  $i \notin E$ , it follows that

$$P_i(\sigma > t(\varepsilon)) < 2\delta$$

for all sufficiently small  $\varepsilon$ . Notice also that for  $i \in E$ ,

$$P_i(t_1^\varepsilon \leq t(\varepsilon)) < \delta$$

for all sufficiently small  $\varepsilon$ , as follows from the condition that  $t(\varepsilon) \ll T(i, \varepsilon)$  for  $i \in E$ . Since  $\delta$  was arbitrary, using the last two inequalities, (17), and the strong Markov property of the process, we obtain (18).  $\square$

**Remark.** The quantity  $P_i(X_\sigma^\varepsilon = j)$  can be represented in terms of  $i$ -graphs (see Chapter 6 of [5]). Such a representation could be used as an alternative way to prove Lemma 3.3.

## 4 Metastable distributions for completely asymptotically regular families.

### 4.1 Formulation of the result.

Suppose that  $X_t^\varepsilon$  is completely asymptotically regular. In this section, we'll study the distribution of  $X_{t(\varepsilon)}^\varepsilon$  at time scales  $t(\varepsilon)$  that vary with  $\varepsilon$  as  $\varepsilon \downarrow 0$ . The initial state  $X_0^\varepsilon$  is assumed to be fixed.

To give a clearer picture, we first formulate the result in a particular case, with the general case to follow.

**Theorem 4.1.** *Let  $X_t^\varepsilon$  be asymptotically regular and  $N \geq 2$ . Suppose that the skeleton chain has one ergodic class and no transient states. Suppose that for each  $1 \leq i \leq N$  either  $t(\varepsilon) \ll T(i, \varepsilon)$  or  $t(\varepsilon) \gg T(i, \varepsilon)$ . Then there is a family of probability measures  $\nu(i, \cdot)$ ,  $i \in S$ , on  $S$  such that*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = \nu(i, j).$$

The measure  $\nu(i, \cdot)$  will be referred to as the metastable distribution for the initial state  $i$  at the time scale  $t(\varepsilon)$ .

*Proof.* Let  $E = \{i \in S : t(\varepsilon) \ll T(i, \varepsilon)\}$ . If  $E = \emptyset$ , then the result follows from (9), i.e., the limiting measure is the invariant measure, for each initial state. If  $E \neq \emptyset$ , then the result follows from (18), i.e., the limiting measure is concentrated on  $E$  and may depend on the initial point  $i$ .  $\square$

Now let formulate the result in the general case and describe how to identify the metastable distributions. For  $0 \leq r \leq \rho$ , the *inverse transition rate* of  $S_i^r$  is defined as

$$T^r(i, \varepsilon) = \left( \sum_{j \neq i} Q_{ij}^r(\varepsilon) \right)^{-1}, \quad (19)$$

where the sum is over  $1 \leq j \leq n_r$ ,  $j \neq i$  (if  $S_i^r \in S_k^{r+1}$  and  $S_k^{r+1} \neq \{S_i^r\}$ , this is asymptotically equivalent to taking only such  $j$  that  $S_j^r \in S_k^{r+1}$ ). Note that  $T^r(i, \varepsilon) \equiv +\infty$  if and only if  $r = \rho$ . Our main result is the following.

**Theorem 4.2.** *Suppose that for each  $0 \leq r \leq \rho - 1$  and each  $1 \leq i \leq n_r$  either  $t(\varepsilon) \ll T^r(i, \varepsilon)$  or  $t(\varepsilon) \gg T^r(i, \varepsilon)$ . Then there is a family of probability measures  $\nu(i, \cdot)$ ,  $i \in S$ , on  $S$  such that*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = \nu(i, j).$$

The proof of the theorem will be based on two lemmas that we formulate next. Let

$$\mu^{r,k}(i) = \lim_{\varepsilon \downarrow 0} \mu^{r,k}(i, \varepsilon) \text{ if } S_i^r \in S_k^{r+1}; \quad \mu^{r,k}(i) = 0 \text{ if } S_i^r \notin S_k^{r+1}, \quad (20)$$

where the existence of the limit is guaranteed by Lemma 3.1. For  $-1 \leq r < \rho$  and  $j \prec S_k^{r+1}$  such that  $j = S_{j_0}^0 \in S_{j_1}^1 \dots \in S_{j_{r+1}}^{r+1} = S_k^{r+1}$ , define

$$\nu^r(j) = \mu^{0,j_1}(j_0)\mu^{1,j_2}(j_1)\dots\mu^{r,j_{r+1}}(j_r), \quad (21)$$

where the right hand side is defined to be one if  $r = -1$ . For  $i \in S$ , let  $r(i)$  be the minimal value of  $r$  such that there is  $k$  with  $i \prec S_k^{r+1}$  and  $t(\varepsilon) \ll T^{r+1}(k, \varepsilon)$ . Let  $L(i) = \{l : S_l^{r(i)} \in S_k^{r(i)+1} \text{ and } t(\varepsilon) \ll T^{r(i)}(l, \varepsilon)\}$ . Obviously,  $L(i) = \emptyset$  if  $r(i) = -1$  since then there are no  $l$  for which  $S_l^{r(i)}$  is defined. The following lemma (proved in the next subsection) provides the description of the metastable distribution in some cases.

**Lemma 4.3.** *Let the assumption made in Theorem 4.2 hold. Let  $k$  be such that  $i \prec S_k^{r(i)+1}$ . Suppose that  $L(i) = \emptyset$ . If  $j$  satisfies (5), then*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = \nu^{r(i)}(j). \quad (22)$$

If  $j \prec S_k^{r(i)+1}$  does not hold, then  $\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = 0$ .

Now consider the case when  $L = L(i) \neq \emptyset$ . Let  $l(i)$  be such that  $i \prec S_{l(i)}^r$ , where  $r = r(i)$ . Observe that  $l(i) \notin L(i)$ . Recall that the state space of  $Y_t^{r,k,\varepsilon}$  is  $S_k^{r+1}$ . We define a new state space  $\tilde{S}_k^{r+1}$  by removing all the states  $S_l^r$ ,  $l \in L$ , from  $S_k^{r+1}$  and adjoining the set  $E = \{j : j \prec S_l^r \text{ for some } l \in L\}$ . On this new state space, we define the Markov chain  $\tilde{Y}_t^\varepsilon$ . Its transition rates are defined as follows. If  $l, m \notin L$ ,  $l \neq m$ , are such that  $S_l^r, S_m^r \in S_k^{r+1}$ , then

$$\tilde{Q}_{S_l^r S_m^r}(\varepsilon) = Q_{lm}^r(\varepsilon).$$

The transition rates between  $S_l^r \subset S_k^{r+1}$ ,  $l \notin L$ , and  $j \in E$  are defined as

$$\tilde{Q}_{S_l^r j}(\varepsilon) = \tilde{Q}_{lj}^r(\varepsilon),$$

where the quantity in the right hand side has been defined after the construction of the hierarchy. The transition rate from  $j \in E$  to any other state is zero, i.e.,  $E$  is the terminal set. It is not difficult to show that  $\tilde{Y}_t^\varepsilon$  satisfies the assumptions of Lemma 3.3, i.e., it coincides with an asymptotically regular Markov chain stopped upon entering  $E$ . (The proof of this statement is the same as the proof of Lemma 2.1.) Therefore, by (17), there is a probability measure  $\eta(i, \cdot)$  on  $E$  such that

$$\lim_{\varepsilon \downarrow 0} \lim_{t \rightarrow \infty} P(\tilde{Y}_t^\varepsilon = j | \tilde{Y}_0^\varepsilon = S_{l(i)}^r) = \eta(i, j), \quad j \in E.$$

The proof of the following lemma is similar to that of Lemma 4.3, and therefore not presented here.

**Lemma 4.4.** *Let the assumption made in Theorem 4.2 hold. Suppose that  $j \prec S_m^{r(i)}$ , where  $m \in L(i)$ . Then*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = \sum_{i' \prec S_m^{r(i)}} \eta(i, i') \lim_{\varepsilon \downarrow 0} P_{i'}(X_{t(\varepsilon)}^\varepsilon = j),$$

provided that the limits in the right hand side exist. If  $j$  is such that  $j \prec S_m^{r(i)}$  does not hold for any  $m \in L$ , then

$$\lim_{\varepsilon \downarrow 0} P_i(X_{t(\varepsilon)}^\varepsilon = j) = 0.$$

## 4.2 Proof of the main result.

It is clear that Lemmas 4.3 and 4.4 imply Theorem 4.2. Indeed, if  $L(i) = \emptyset$ , then the metastable distribution is given by Lemma 4.3. If  $L(i) \neq \emptyset$ , then, by Lemma 4.4,  $\nu(i, j)$  is either equal to zero or is equal to a linear combination of the quantities  $\nu(i', j)$  with  $r(i') < r(i)$  (provided that all  $\nu(i', j)$  are defined). The values of  $\nu(i', j)$  can be found from Lemma 4.3 (when  $L(i') = \emptyset$ ), or again expressed in terms of metastable distributions with different initial points, using Lemma 4.4. This recursive procedure can be continued until all the resulting initial points  $j$  satisfy  $L(j) = \emptyset$ , in which case Lemma 4.3 can be applied (which will happen in no more than  $r(i) + 1$  steps since  $L(j) = \emptyset$  whenever  $r(j) = -1$ ).

It remains to prove Lemmas 4.3 and 4.4. We give the proof of Lemmas 4.3 and omit the proof of Lemma 4.4, since it is quite similar.

*Proof of Lemma 4.3.* Let us start by briefly explaining the main idea of the proof. First, let us “reduce” the state space of  $X_t^\varepsilon$  by clumping all the states  $i$  with the property that  $i \prec S_m^{r(i)}$  into a single state (recall the definition of ' $\prec$ ' from (5)). The resulting process is well-approximated by the Markov chain  $Y_t^{r(\varepsilon), k, \varepsilon}$ , where  $k$  is such that  $X_0^\varepsilon \prec S_k^{r(i)+1}$ . Let us observe the process on a time scale  $s(\varepsilon) \sim t(\varepsilon)$ , such that  $s(\varepsilon)$  is slightly smaller than  $t(\varepsilon)$ . Take  $m$  such that  $j \prec S_m^{r(i)}$ . From the properties of  $Y_t^{r(\varepsilon), k, \varepsilon}$  it then follows that  $\lim_{\varepsilon \downarrow 0} P_i(X_{s(\varepsilon)}^\varepsilon \prec S_m^{r(i)}) = \mu^{r(i), k}(m)$ , yielding the last factor in the expression (21). Next, we can consider the process  $X_t^\varepsilon$  on the time scale  $t(\varepsilon) - s(\varepsilon)$  starting at a point  $i' = X_{s(\varepsilon)}^\varepsilon \prec S_m^{r(i)}$ . Provided that  $s(\varepsilon)$  is chosen appropriately, the problem of identifying the limiting distribution at this time scale is similar to the original one, but with  $r(i') = r(i) - 1$ . Iterating the argument  $r(i)$  times, we'll get the desired distribution. Let us now make the above arguments formal.

For  $0 \leq r < \rho$ , define the process  $\bar{Y}_t^{r, \varepsilon}$  via

$$\bar{Y}_t^{r, \varepsilon} = S_m^r \quad \text{if } X_t^\varepsilon \prec S_m^r.$$

This is not necessarily a Markov process. However, below we'll see that on certain time scales it is close, in a certain sense, to the Markov process  $Y_t^{r, k, \varepsilon}$ , where  $k$  is such that  $X_0^\varepsilon = i \prec S_k^{r+1}$ . Suppose that  $s(\varepsilon)$  is such that  $s(\varepsilon) \ll T^{r+1}(k, \varepsilon)$ , yet  $T^r(l, \varepsilon) \ll s(\varepsilon)$  whenever  $S_l^r \in S_k^{r+1}$ . We claim (and will prove below) that

$$\lim_{\varepsilon \downarrow 0} P(\bar{Y}_{s(\varepsilon)}^{r, \varepsilon} = S_l^r | X_0^\varepsilon = i) = \mu^{r, k}(l) \quad (23)$$

for each  $1 \leq l \leq n_r$ , where  $\mu^{r, k}$ , defined in (20), is the limit, as  $\varepsilon \downarrow 0$ , for the invariant measures of the processes  $Y_t^{r, k, \varepsilon}$ .

Recall that  $t(\varepsilon) \ll T^{r(i)+1}(k, \varepsilon)$ , while  $T^{r(i)}(l, \varepsilon) \ll t(\varepsilon)$  whenever  $S_l^{r(i)} \in S_k^{r(i)+1}$ . Choose  $s(\varepsilon) < t(\varepsilon)$  with the following properties:

- (a)  $s(\varepsilon) \ll T^{r(i)+1}(k, \varepsilon)$ , while  $T^{r(i)}(l, \varepsilon) \ll s(\varepsilon)$  whenever  $S_l^{r(i)} \in S_k^{r(i)+1}$ .
- (b) For each  $l$  such that  $\mu^{r(i),k}(l) > 0$ , we have  $t(\varepsilon) - s(\varepsilon) \ll T^{r(i)}(l, \varepsilon)$ , while  $T^{r'}(l', \varepsilon) \ll t(\varepsilon) - s(\varepsilon)$  whenever  $S_{l'}^{r'}$  is such that  $S_{l'}^{r'} \prec S_l^{r(i)}$  and there is  $j \prec S_l^{r(i)}$  for which  $j \prec S_{l'}^{r'}$  does not hold. (Here  $S_{l'}^{r'} \prec S_l^{r(i)}$  means that  $S_{l'}^{r'} \in S_{l_1}^{r'+1} \in \dots \in S_{l_m}^{r'+m} \in S_l^{r(i)}$  for some  $m, l_1, \dots, l_m$ .)

The existence of such  $s(\varepsilon)$  follows from Lemma 3.1, which guarantees that the quantities  $T^{r(i)}(l, \varepsilon)$  (with  $l$  such that  $\mu^{r(i),k}(l) > 0$ ) are all asymptotically equivalent, up to multiplicative constants.

By property (a), from (23) with  $r = r(i)$ , it follows that  $X_{s(\varepsilon)}^\varepsilon \prec S_l^{r(i)}$  with probability close to  $\mu^{r(i),k}(l)$  for  $l$  such that  $\mu^{r(i),k}(l) > 0$ . Then, by the Markov property,  $X_{s(\varepsilon)}^\varepsilon$  can be taken as a new starting point for the process that will be studied on the time interval  $t(\varepsilon) - s(\varepsilon)$ . For this new time scale and the new initial point, all the assumptions of Lemma 4.3 are satisfied, with the exception that the value of  $r(i)$  is reduced by at least one. Thus, conditioning on the event that  $\bar{Y}_{s(\varepsilon)}^{r(i),\varepsilon} = S_l^r$ , with  $l$  such that  $j \prec S_l^{r(i)}$ , gives

$$\lim_{\varepsilon \downarrow 0} \mathbb{P}_i(X_{t(\varepsilon)}^\varepsilon = j) = \mu^{r(i),k}(l) \lim_{\varepsilon \downarrow 0} \mathbb{P}_{i'}(X_{t(\varepsilon)-s(\varepsilon)}^\varepsilon = j),$$

provided that the limit in the right hand side is defined and does not depend on  $i' \prec S_l^{r(i)}$ . Iterating this argument  $r(i) + 1$  times gives (22). It remains to prove (23).

For  $i \in S$  and  $-1 \leq r < \rho$ , let  $k$  be such that  $i \prec S_k^{r+1}$ . Let  $A^r = \{j : j \prec S_k^{r+1}\}$ . For  $0 \leq r < \rho$ , let  $\tilde{X}_t^{r,\varepsilon}$  be the Markov chain on the state space  $A^r$ , whose transition rates agree with those of the chain  $X_t^\varepsilon$  (i.e.,  $\tilde{X}_t^{r,\varepsilon}$  is obtained from  $X_t^\varepsilon$  by disallowing transitions outside the specified state space).

For  $j \prec S_k^{r+1}$  such that  $j = S_{j_0}^0 \in S_{j_1}^1 \dots \in S_{j_{r+1}}^{r+1}$  with  $j = j_0, k = j_{r+1}$ , let

$$M^r(j, \varepsilon) = \mu^{0,j_1}(j_0, \varepsilon) \mu^{1,j_2}(j_1, \varepsilon) \dots \mu^{r,j_{r+1}}(j_r, \varepsilon).$$

For  $0 \leq r < \rho$ , let  $\alpha^r(j, s(\varepsilon))$  be the proportion of time, prior to  $s(\varepsilon)$ , that the process  $\tilde{X}_t^{r,\varepsilon}$  spends in  $j$ . Let

$$\bar{T}^r(\varepsilon) = \sum_{l: S_l^r \in S_k^{r+1}} \mu^{r,k}(l) T^r(l, \varepsilon).$$

Define the process  $\tilde{Y}_t^{r,\varepsilon}$  via

$$\tilde{Y}_t^{r,\varepsilon} = S_m^r \quad \text{if} \quad \tilde{X}_t^{r,\varepsilon} \prec S_m^r.$$

Recall that  $s(\varepsilon)$  satisfies  $s(\varepsilon) \ll T^{r+1}(k, \varepsilon)$  and  $T^r(l, \varepsilon) \ll s(\varepsilon)$  whenever  $S_l^r \in S_k^{r+1}$ . We claim that for each  $0 \leq r < \rho$  and  $s(\varepsilon)$  satisfying these conditions, the following asymptotic relations hold. For each  $i, j \in A^r$  and  $l$  such that  $S_l^r \in S_k^{r+1}$ ,

$$\lim_{\varepsilon \downarrow 0} \mathbb{P}(\tilde{Y}_{s(\varepsilon)}^{r,\varepsilon} = S_l^r | \tilde{X}_0^{r,\varepsilon} = i) = \mu^{r,k}(l). \quad (24)$$

For each  $c > 0$ , there are  $\delta(c) > 0$  and a function  $\varphi(\varepsilon) > 0$  such that  $\lim_{\varepsilon \downarrow 0} \varphi(\varepsilon) = +\infty$  and

$$P_i(\alpha^r(j, s(\varepsilon)) < (1 - c)M^r(j, \varepsilon)) \leq e^{-\delta(c)\varphi(\varepsilon)}, \quad (25)$$

$$P_i(\alpha^r(j, s(\varepsilon)) > (1 + c)M^r(j, \varepsilon)) \leq e^{-\delta(c)\varphi(\varepsilon)}. \quad (26)$$

When  $r = 0$ , (24)-(26) follow from Lemma 3.2. Next, let us sketch the inductive step, i.e., the proof that relations (24)-(26) hold for a fixed  $r > 0$ , assuming that they hold for all the smaller values of  $r$ .

The process  $\tilde{Y}_t^{r,\varepsilon}$  is close to the Markov process  $Y_t^{r,k,\varepsilon}$  in the following sense.

(a) If  $l \neq m$  and  $S_l^r, S_m^r \in S_k^{r+1}$ ,  $i \prec S_l^r$ , then

$$P(\text{the first jump of } \tilde{Y}_t^{r,\varepsilon} \text{ is to } S_m^r | \tilde{X}_0^{r,\varepsilon} = i) \sim Q_{lm}^r(\varepsilon) \text{ as } \varepsilon \downarrow 0.$$

(b) If  $\beta^\varepsilon$  is the random time till the first transition of  $\tilde{Y}_t^{r,\varepsilon}$ , then

$$E(\beta^\varepsilon | \tilde{X}_0^{r,\varepsilon} = i) \sim T^r(l, \varepsilon), \text{ as } \varepsilon \downarrow 0,$$

and there are  $\delta > 0$  and  $\varepsilon_0 > 0$  such that

$$P(\beta^\varepsilon \geq \lambda T^r(l, \varepsilon) | \tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta\lambda}, \quad \lambda \geq 2, \quad \varepsilon \in (0, \varepsilon_0].$$

The validity of (a) and (b) easily follows by examining the Markov chain  $\tilde{X}_t^{r-1,\varepsilon}$  and utilizing the fact that (25)-(26) hold with  $r$  replaced by  $r - 1$ .

Lemma 3.2 (with  $s(\varepsilon)$  instead of  $t(\varepsilon)$ ) can be applied to the process  $Y_t^{r,k,\varepsilon}$ . However, we are interested a similar result for the process  $\tilde{Y}_t^{r,\varepsilon}$ . It is not difficult to see that conditions (a) and (b) on the transition probabilities and transition times are sufficient for the proof of Lemma 3.2 to go through and for the result to be valid for the process  $\tilde{Y}_t^{r,\varepsilon}$ . Thus we have (24). Moreover, for  $i \prec S_l^r$  and each  $c > 0$ , there are  $\delta(c) > 0$  and a function  $\varphi(\varepsilon) > 0$  such that  $\lim_{\varepsilon \downarrow 0} \varphi(\varepsilon) = +\infty$  and

$$P(\tilde{\alpha}^r(l, s(\varepsilon)) < (1 - c)\mu^{r,k}(l, \varepsilon) | \tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta(c)\varphi(\varepsilon)},$$

$$P(\tilde{\alpha}^r(l, s(\varepsilon)) > (1 + c)\mu^{r,k}(l, \varepsilon) | \tilde{X}_0^{r,\varepsilon} = i) \leq e^{-\delta(c)\varphi(\varepsilon)},$$

where  $\tilde{\alpha}^r(l, s(\varepsilon))$  is the proportion of time, prior to  $s(\varepsilon)$ , that the process  $\tilde{Y}_t^{r,\varepsilon}$  spends in the state  $S_l^r$ . Together with (25)-(26) for  $r - 1$  instead of  $r$ , these are easily seen to control the proportion of time, prior to  $s(\varepsilon)$ , that  $\tilde{X}_t^{r,\varepsilon}$  spends in  $j$ , thus yielding (25)-(26) for  $r$ .

From (25)-(26) it easily follows that

$$P_i(X_t^\varepsilon \prec S_k^{r+1} \text{ for all } t \leq s(\varepsilon)) \rightarrow 1 \text{ as } \varepsilon \downarrow 0.$$

Thus (23) follows from (24). □

## 5 Complete asymptotic regularity

In this section, we prove Lemma 2.1 and briefly discuss a couple of generalizations. We start with the following simple lemma.

**Lemma 5.1.** *Suppose that  $a_1(\varepsilon), \dots, a_n(\varepsilon)$  and  $b_1(\varepsilon), \dots, b_{n'}(\varepsilon)$  are positive functions such that there are limits*

$$\lim_{\varepsilon \downarrow 0} \frac{a_i(\varepsilon)}{b_{i'}(\varepsilon)} \in [0, \infty], \quad 1 \leq i \leq n, \quad 1 \leq i' \leq n'.$$

Let  $A(\varepsilon) = a_1(\varepsilon) + \dots + a_n(\varepsilon)$ ,  $B(\varepsilon) = b_1(\varepsilon) + \dots + b_{n'}(\varepsilon)$ . Then there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{B(\varepsilon)} \in [0, \infty].$$

*Proof.* For each  $i'$ , there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{b_{i'}(\varepsilon)} = \lim_{\varepsilon \downarrow 0} \frac{a_1(\varepsilon)}{b_{i'}(\varepsilon)} + \dots + \lim_{\varepsilon \downarrow 0} \frac{a_n(\varepsilon)}{b_{i'}(\varepsilon)} \in [0, \infty].$$

Therefore, there is the limit

$$\lim_{\varepsilon \downarrow 0} \frac{A(\varepsilon)}{B(\varepsilon)} = \left( \lim_{\varepsilon \downarrow 0} \frac{b_1(\varepsilon)}{A(\varepsilon)} + \dots + \lim_{\varepsilon \downarrow 0} \frac{b_{n'}(\varepsilon)}{A(\varepsilon)} \right)^{-1} \in [0, \infty].$$

□

*Proof of Lemma 2.1.* The case  $N = 1$  is trivial. Let us assume that  $N \geq 2$ . Recall the decomposition (2) of  $S$  into ergodic classes and transient states. To prove the lemma, we need to show that for  $a \geq 1$  there is the limit

$$\lim_{\varepsilon \downarrow 0} \left( \frac{Q_{k_1 l_1}(\varepsilon)}{Q_{m_1 n_1}(\varepsilon)} \times \dots \times \frac{Q_{k_a l_a}(\varepsilon)}{Q_{m_a n_a}(\varepsilon)} \right) \in [0, \infty], \quad (27)$$

provided that  $k_1 \neq l_1, \dots, k_a \neq l_a, m_1 \neq n_1, \dots, m_a \neq n_a$ . We will repeatedly use Lemma 5.1, which will allow us to replace each of the factors above by simpler expressions. First consider a factor of the form  $Q_{kl}/Q_{mn}$  under the assumption that  $S_k = \{i\}$  and  $S_m = \{i'\}$ , i.e.,  $S_k$  and  $S_m$  have only one element each. In this case

$$\frac{Q_{kl}(\varepsilon)}{Q_{mn}(\varepsilon)} = \frac{\sum_{j \in S_l} q_{ij}(\varepsilon)}{\sum_{j' \in S_n} q_{i'j'}(\varepsilon)}.$$

Thus, by Lemma 5.1, it is sufficient to prove the existence of the limit in (27) with all such factors replaced by those of the form  $q_{ij}/q_{i'j'}$ .

Next consider a factor of the form  $Q_{kl}/Q_{mn}$  under the assumption that one of the sets  $S_k$  and  $S_m$  (say,  $S_k$ ) has at least two elements, while the other one has one element, i.e.,

$S_m = \{i'\}$ . Then the chain  $Y_t^{k,\varepsilon}$  satisfies the assumptions of Lemma 3.1, and therefore, by (4),

$$\frac{Q_{kl}(\varepsilon)}{Q_{mn}(\varepsilon)} = \frac{\sum_{i \in S_k} \sum_{j \in S_l} \mu^k(i, \varepsilon) q_{ij}(\varepsilon)}{\sum_{j' \in S_n} q_{i'j'}(\varepsilon)} \sim \frac{\sum_{i \in S_k} \sum_{j \in S_l} \lambda(i) T(i, \varepsilon) q_{ij}(\varepsilon)}{\sum_{j' \in S_n} q_{i'j'}(\varepsilon) \sum_{i'' \in S_k} (\lambda(i'') T(i'', \varepsilon))},$$

where  $\lambda$  and  $T$  are the invariant measure for the skeleton chain and the inverse transition rate, respectively, for the chain  $Y_t^{k,\varepsilon}$ . Thus, by Lemma 5.1, it is sufficient to prove the existence of the limit in (27) with  $Q_{kl}/Q_{mn}$  replaced by  $T(i, \varepsilon) q_{ij}(\varepsilon) (T(i'', \varepsilon) q_{i'j'}(\varepsilon))^{-1}$ , where  $i, i'' \in S_k$ . By the definition of  $T$ ,

$$\frac{T(i, \varepsilon) q_{ij}(\varepsilon)}{T(i'', \varepsilon) q_{i'j'}(\varepsilon)} = \frac{\sum_{b' \in S_k, b' \neq i''} q_{i''b'}(\varepsilon) q_{ij}(\varepsilon)}{\sum_{b \in S_k, b \neq i} q_{ib}(\varepsilon) q_{i'j'}(\varepsilon)}.$$

By Lemma 5.1, each such expression can be replaced by  $(q_{i''b'}(\varepsilon) q_{ij}(\varepsilon)) / (q_{ib}(\varepsilon) q_{i'j'}(\varepsilon))$ .

The final case, when  $S_k$  and  $S_m$  have at least two elements each, is treated similarly, resulting in  $Q_{kl}/Q_{mn}$  being replaced by a product of three factors of the form  $q_{ij}/q_{i'j'}$ . Thus we see that each factor in (27) can be replaced by either one, two, or three factors of the form  $q_{ij}/q_{i'j'}$ . Therefore, the limit in (27) exists since the original chain is completely asymptotically regular.  $\square$

## 6 Remarks and generalizations.

(A) One could replace the complete asymptotic regularity by a somewhat weaker assumption that also implies the asymptotic regularity of all the reduced chains appearing in the inductive construction of the hierarchy. We will say that an asymptotically regular family of Markov chains  $X_t^\varepsilon$  satisfies Condition (A) if for each  $0 \leq a \leq N - 2$  the following finite or infinite limit exists

$$\lim_{\varepsilon \downarrow 0} \left( \frac{q_{k_1 l_1}(\varepsilon)}{q_{k_1 n_1}(\varepsilon)} \times \dots \times \frac{q_{k_a l_a}(\varepsilon)}{q_{k_a n_a}(\varepsilon)} \times \frac{q_{k' l'}(\varepsilon)}{q_{m' n'}(\varepsilon)} \right) \in [0, \infty],$$

provided that  $k_1, \dots, k_a$  are all distinct,  $k', m' \notin \{k_1, \dots, k_a\}$ , and  $k_1 \neq l_1, \dots, k_a \neq l_a$ ,  $k_1 \neq n_1, \dots, k_a \neq n_a$ ,  $k' \neq l', m' \neq n'$ . The proof of the following lemma is similar to that of Lemma 2.1, and so we don't provide it here.

**Lemma 6.1.** *If  $X_t^\varepsilon$  satisfies Condition (A), then the reduced chain also satisfies Condition (A).*

If  $X_t^\varepsilon$  satisfies Condition (A), then, by Lemma 6.1, so do the reduced Markov chains that appear at each step of the inductive construction of the hierarchy, which implies that all of them are asymptotically regular.

(B) Next, let us mention that all the above analysis can be easily adapted to the case



of discrete-time Markov chains. Instead of (1), we assume that  $X_n^\varepsilon$ ,  $n \geq 0$ , is a Markov chain with transition probabilities  $p_{ij}(\varepsilon)$ ,  $i, j \in S$ . We impose an additional assumption that  $p_{ii}(\varepsilon) \geq c > 0$  for all  $i \in S$ ,  $\varepsilon > 0$ . This is needed in order for the discrete time analogue of Lemma 3.2 to remain valid.

The definitions of asymptotic regularity and complete asymptotic regularity remain the same as in the continuous time case, with transition rates  $q_{ij}(\varepsilon)$  replaced by transition probabilities  $p_{ij}(\varepsilon)$ . The reduced Markov chains the chains  $Y_t^{r,k,\varepsilon}$  can be still defined in continuous time, simply replacing  $q_{ij}(\varepsilon)$  by  $p_{ij}(\varepsilon)$  in all the definitions. The definition of the inverse transition rates (19) remains the same. The theorem on metastable distributions now takes the following form.

**Theorem 6.2.** *Let  $n : (0, \infty) \rightarrow \mathbb{N}$  be such that for each  $0 \leq r \leq \rho - 1$  and each  $1 \leq i \leq n_r$  either  $n(\varepsilon) \ll T^r(i, \varepsilon)$  or  $n(\varepsilon) \gg T^r(i, \varepsilon)$ . Then there is a family of probability measures  $\nu(i, \cdot)$ ,  $i \in S$ , on  $S$  such that*

$$\lim_{\varepsilon \downarrow 0} P_i(X_{n(\varepsilon)}^\varepsilon = j) = \nu(i, j).$$

The proof of this theorem is identical to that of Theorem 4.2.

(C) Finally, consider an example of a completely asymptotically regular family of Markov chains. Given numbers  $\alpha_{ij}$ ,  $\beta_{ij}$ , and  $\gamma_{ij}$ ,  $i \neq j$ , such that  $\alpha_{ij} > 0$ , assume that the transition rates satisfy

$$q_{ij}(\varepsilon) \sim \alpha_{ij} \varepsilon^{\beta_{ij}} \exp(-\gamma_{ij} \varepsilon^{-1}), \quad \text{as } \varepsilon \downarrow 0, \quad i \neq j. \quad (28)$$

It is clear that these functions satisfy (6), and therefore the corresponding family is completely asymptotically regular. Replacing  $\varepsilon$  by  $|\ln \tilde{\varepsilon}|^{-1}$ , we obtain functions

$$\tilde{q}_{ij}(\tilde{\varepsilon}) \sim \alpha_{ij} |\ln \tilde{\varepsilon}|^{-\beta_{ij}} (\tilde{\varepsilon})^{\gamma_{ij}}, \quad \text{as } \varepsilon \downarrow 0, \quad i \neq j, \quad (29)$$

which also satisfy (6). The systems discussed in the Introduction lead to Markov chains with transition rates that satisfy either (28) or (29), with the exception that the condition  $\alpha_{ij} > 0$  may be violated, i.e., some of the coefficients may be equal to zero. In fact, this positivity condition (or condition (a) in the definition of asymptotic regularity) are not that crucial. If  $i$  and  $j$  are such that  $\alpha_{ij} = 0$ , the transition rates can be re-defined for those  $(i, j)$  by taking  $\alpha_{ij} = 1$  and  $\gamma_{ij}$  sufficiently large, resulting in a completely asymptotically regular family with the same metastable behavior as the original one.

**Acknowledgements:** While working on this article, M. Freidlin was supported by NSF grant DMS-1411866 and L. Koralov was supported by NSF grant DMS-1309084.

## References

- [1] Bakhtin Y., Pajor-Gyulai Z., *Metastability and cycle structure in strictly attracting noisy heteroclinic networks*, in preparation.
- [2] Doob J. L., *Stochastic Processes*, Wiley-Interscience; Revised edition (January 25, 1990).
- [3] Freidlin M. I., *Sublimiting distributions and stabilization of solutions of parabolic equations with a small parameter*, Soviet Math. Dokl., 18, No 4 (1977), 1114–1118.
- [4] Freidlin M. I., *On Stochastic Perturbations of Systems with Rough Symmetry. Hierarchy of Markov Chains*, Journal of Statistical Physics, Vol. 157, No 6, pp 1031-1045, 2014.
- [5] Freidlin M. I., Wentzell A. D., *Random Perturbations of Dynamical Systems*, third edition, Springer 2012.
- [6] Freidlin M. I., Koralov L., Wentzell A. D., *On the behavior of diffusion processes with traps*, to appear in Annals of Probability.
- [7] Holmes-Cerfon M., Gortler S. J., Brenner M. P., *A geometrical approach to computing free-energy landscapes from short-ranged potentials*, Proc. Natl. Acad. Sci., 110 (1), pp E5-E14, 2013.
- [8] Manzo F., Nardi F. R., Olivieri E., Scoppola E., *On the Essential Features of Metastability: Tunnelling Time and Critical Configurations*, Journal of Statistical Physics, Vol. 115, Nos. 1/2, April 2004.
- [9] Olivieri E., Scoppola E., *Markov Chains with Exponentially Small Transition Probabilities: First Exit Problem from a General Domain. I. The Reversible Case*, Journal of Statistical Physics, Vol. 79, Nos. 3/4, 1995.
- [10] Olivieri E., Scoppola E., *Markov Chains with Exponentially Small Transition Probabilities: First Exit Problem from a General Domain. II. The General Case*, Journal of Statistical Physics, Vol. 84, Nos. 5/6, 1996.
- [11] Wentzell A. D., *On the asymptotics of eigenvalues of matrices with elements of order  $\exp(-V_{ij}/(2\varepsilon^2))$* , Soviet Math. Dokl., 13, No 1 (1972), 65-68.